AUDIOVISUAL SPEECH PERCEPTION IN PEOPLE WITH HEARING LOSS ACROSS LANGUAGES: A SYSTEMATIC REVIEW OF ENGLISH AND MANDARIN

by

Ya-Wen Tsao

A Thesis Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by ***Ya-Wen Tsao***, titled ***Audiovisual Speech Perception in People With Hearing Loss Across Languages: A Systematic Review of English and Mandarin*** and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

Nicole Marrone, Ph.D.
Dr. Nicole Marrone

Date: ***February 4, 2019***

Dr. Leah Fabiano-Smith

Date: ***February 4, 2019***

Dr. Brad Story

Date: ***February 4, 2019***

Aileen Wong, Au.D.
Dr. Aileen Wong

Date: ***February 4, 2019***

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

Nicole Marrone
Dr. Nicole Marrone
Associate Professor
Speech, Language, and Hearing Sciences

Date: ***February 4, 2019***

Acknowledgments

I would like to express my gratitude to Dr. Nicole Marrone for guiding me through and for helping me with the thesis. I would like to thank Dr. Marrone for reviewing the titles, abstracts, and the articles with the full-text, for assessing the level of evidence and quality of the articles included in the systematic review, and for assessing the confidence in cumulative evidence. I would also like to thank Dr. Leah Fabiano-Smith, Dr. Brad Story, and Dr. Aileen Wong for providing me with valuable suggestions for the thesis. I would like to thank Dr. Alyssa Everett for helping me with the systematic review and for reviewing the abstracts. I would also like to thank Sandra Kramer for assisting me in the database searching and in the search strategies.

I am deeply grateful to my family who have always been there supporting me.

Table of Contents

List of Figures and Tables

List of Figures

List of Tables

Abstract

Audiovisual (AV) information has been reported to facilitate speech understanding among the English-speaking population. However, it is not clear whether audiovisual benefits also exist among people who speak languages other than English. A systematic review was conducted to investigate the audiovisual effects on speech perception among people with hearing loss who speak English and people with hearing loss who speak Mandarin. The results of the review demonstrated audiovisual benefits in the English-speaking population with hearing loss regardless of age, degree of hearing loss, use and type of hearing technology, and acoustic environment. By contrast, significant audiovisual benefits were only found for Mandarin phoneme and word recognition but not for tone recognition in pre-lingually deafened adults with cochlear implants and for phoneme recognition in children with hearing aids. No significant audiovisual benefits were revealed in Mandarin-speaking post-lingually deafened adults with cochlear implants and for speech perception at higher intensity levels. Heterogeneity in the results across studies and limitations of the included studies were discussed.

Keywords: audiovisual, speech perception, hearing loss, English, Mandarin, systematic review

Introduction

According to the World Health Organization (2018), more than 5% of the world population has a hearing loss that exceeds the definition of disabling hearing loss[1]. In the United States, auditory habilitation and rehabilitation often promote utilizing visual cues to help communication (American Speech-Language-Hearing Association, n.d.-a, n.d.-b). Research evidence shows benefits of visual information for speech perception in English for both individuals with normal hearing and those with hearing loss. Visual information has been reported to improve speech intelligibility when the signal-to-noise ratio decreases among normal hearing individuals for English bisyllabic word stimuli (Sumby & Pollack, 1954). Grant, Walden, and Seitz (1998) reported that sentence recognition at a signal-to-noise ratio of 0 dB was better in the audiovisual condition than in the auditory-only (A) condition in American English-speaking participants with hearing loss. Tyler et al. (1997) reported better speech recognition scores in audiovisual conditions than in the auditory-only and visual-only (V) conditions in pre-lingually deaf children with cochlear implants. Audiovisual benefits of speech perception have been shown in the English-speaking population with and without the presence of hearing loss.

The most widely spoken language in the United States is English, with 231 million people who speak only English at home (United States Census Bureau, 2015). However, many other languages are spoken in the U.S. and worldwide. Over 60 million people in the U.S. speak a language other than English at home, with an estimated 37.5 million Spanish speakers and

[1] Disabling hearing loss is defined as a hearing loss in the better ear greater than 40 dB HL and 30 dB HL in adults and in children respectively.

nearly 3 million Chinese speakers (United States Census Bureau, 2015). Globally, English has the third most first-language speakers, following Chinese and Spanish (Simons & Fennig, 2018). These languages have different acoustic and phonological characteristics. Specifically, Chinese is a tonal language while Spanish and English are non-tonal.

For tonal languages, such as Mandarin Chinese, one syllable spoken with different pitch contours can yield different meanings and the evidence has been mixed as to whether visual information is beneficial to speech perception. Liu et al. (2014) reported that word recognition was significantly better in the audiovisual condition than in the auditory-only condition in Mandarin-speaking pre-lingually deafened participants with cochlear implants and participants with normal hearing. It was also found that Mandarin phoneme recognition, but not tone recognition, was significantly better in the audiovisual condition than in the auditory-only condition in the pre-lingually deafened cochlear-implanted and normal-hearing Mandarin-speaking participants. Additionally, Sekiyama (1997) discovered that the McGurk effect (the illusion that occurs when there is a mismatch between auditory and visual stimuli in terms of place of articulation) was significantly weaker in the Chinese-speaking normal hearing participants than in the American English-speaking normal hearing participants. Based on the findings, the author further suggested that the Chinese participants may have a stronger reliance on auditory information and be less susceptible to visual information.

Comparison of English and Mandarin Phonemic Inventories

As background to this systematic review, a comparison of the phonemic inventories of English and Mandarin is provided. The purpose of comparing phonemic inventories is to consider whether there are differences in the numbers of phonemes that are visible in English and Mandarin. Vowels (resonated phonemes) and consonants (articulated phonemes) are

compared separately because of their articulatory and acoustic differences. With respect to vowels, lip rounding is a visible feature of vowels. In English, /u/, /ʊ/, /o/, and /ɔ/ are rounded monophthongs. In Mandarin, rounded monophthongs included /y/ and /u/. Comparing English and Mandarin consonant inventories (see Table 1 & Table 2), there are more phonemes with labial features in English than in Mandarin consonant inventories and equivalent numbers of coronal consonants in English and Mandarin consonant inventories. Phonemes with labial features are easier to be visually observed through lipreading compared with coronal phonemes.

Comparison of English and Mandarin Visemes[2]

In English, phonemes can be grouped into 12 groups of visemes, including consonantal labial, alveolar, velar, labiodental, palato-alveolar, and dental, and vocalic spread, open-spread, neutral, rounded, and protruding-rounded, and silence closed (Jachimski, Czyzewski, & Ciszewski, 2018). While in Mandarin, phonemes are grouped differently and can be grouped into 13 types of visemes, including seven initial types and six finals types (Li & Tang, 2011). The phonemes that are categorized as the same group of visemes in English are not necessarily categorized as the same type of visemes in Mandarin. In addition, the number of phonemes included in an English viseme group may be different from the number of phonemes in a Mandarin viseme group. For example, /f/ and /v/ are the two phonemes of labiodental visemes in English while /f/ is the only one phoneme of Type F viseme in Mandarin.

Visemes may be grouped differently across studies. In terms of consonant visemes, Chen and Massaro (2011) proposed that Mandarin consonants be grouped into eight viseme categories.

[2] Visemes are groups of phonemes that are visually alike.

The authors grouped /d, t, n, l/ into the same viseme group and mentioned that /d, t, n, l/ can be articulated with a visible dental-alveolar tongue in Mandarin while they are alveolar phonemes in English. Taking the articulation differences in /d, t, n, l/ between English and Mandarin into consideration, there are more consonant viseme groups phonemes that are clearly visible in Mandarin than in English. That is, there are three clearly visible consonant viseme groups (seven phonemes) in Mandarin and two visible consonant viseme groups (five phonemes) in English for labial, labiodental and dental-alveolar sounds.

In summary, there are a number of important differences in how phonemes and visemes are represented across English and Mandarin. Different places of articulation may contribute to different degrees of visibility of sounds. Consequently, it is expected that there may also be differences in audiovisual benefits across languages.

Table 1

English Phonemic Inventory

	Bilabial		Labiodental		Dental		Alveolar		Postalveolar		Palatal		Velar		Uvular	Glottal
Plosive	p	b					t	d					k	g		
Nasal		m						n						ŋ		
Trill																
Tap or Flap																
Fricative			f	v	θ	ð	s	z			ʃ	ʒ				h
Affricate									tʃ	dʒ						
Glides		w						r				j				
Liquid								l								

Note. Reprinted from American Speech-Language-Hearing Association (n.d.-c).

Table 2

Mandarin Phonemic Inventory

	Bilabial		Labiodental	Dental	Alveolar		Alveopalatal		Postalveolar	Retroflex		Palatal	Velar		Uvular	Glottal
Plosive	p	p^h			t	t^h							k	k^h		
Nasal		m				n								ŋ		
Trill																
Tap or Flap																
Fricative			f		s		ɕ			ʂ			x			
Affricate					ts	ts^h	tɕ	$tɕ^h$		tʂ	$tʂ^h$					
Glides										ɹ						
Liquid						l										

Note. Reprinted from American Speech-Language-Hearing Association (n.d.-d).

Purpose

The purpose of this systematic review is to investigate audiovisual speech perception across languages among people with hearing loss, specifically for English and Mandarin Chinese. The results of this review may assist clinicians in developing auditory habilitation/rehabilitation programs for people with hearing loss from different language backgrounds. Ultimately, the results of the review can inform people with hearing loss of evidence-based communication strategies to facilitate speech perception.

Methods

Information Sources

Embase, PubMed, Scopus, CINAHL Plus, Web of Science, and PsycINFO databases were searched for relevant studies from the peer-reviewed literature. The search strategies were developed with the assistance of a librarian and adapted for each database as necessary; the search strategies are listed in Appendix A. Searches were be restricted to human studies using database limiters. Due to the lack of a limiter for "human" studies available in Web of Science, animal studies were eliminated manually in the review of articles.

Eligibility Criteria

The inclusion criteria for the characteristics of each study to be reviewed were specified in terms of Population, Intervention, Comparison, and Outcomes (PICO).

Populations. The population of study was people with hearing loss. Hearing loss was defined as hearing thresholds across five octave frequencies (0.25, 0.5, 1, 2, and 4 kHz) greater than 15 dB HL; 15 dB HL was selected as the criterion of hearing loss in this review in order to include a larger pool of studies.

Intervention. The intervention of the included studies was audiovisual input, specifically, the combination of two sensory modalities of information (auditory and visual) for speech communication.

Comparison. Speech perception in audiovisual conditions was compared with that in audio-only and/or visual-only conditions.

Outcomes. The outcomes included available measures of speech perception in quiet and in noise, including discrimination and recognition tasks for phoneme, tone, word, and sentence materials. Insignificant and adverse audiovisual effects were also collected.

Time frame. Searches were not limited to a specific time frame.

Language. Only articles written in English and Chinese were included for feasibility. It is acknowledged that a bias may arise from the limited selection of the two languages.

Data management

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) were followed for this systematic review and the numbers of studies reviewed at each stage of the systematic process were recorded in a PRISMA flow diagram (Moher et al., 2009). References were identified through the database searching using the search strategies. The references were imported into Endnote X7 and duplicates were automatically discarded by Endnote. Duplicates that were not automatically discarded by Endnote were then manually removed. Books, book sections, serials, theses, reviews, conference papers, conference abstracts, editorials, letter, notes, short surveys, non-human studies, and non-English or non-Mandarin studies were removed. Two reviewers (YT & NM) screened titles and then three reviewers (YT, AE, & NM) screened abstracts of the articles that passed the title screening. Any discrepancy in article selection in this stage was resolved by inclusion of the article for full-text review. After the screening stage, two reviewers (YT & NM) reviewed and assessed the articles with full-text for eligibility. Any discrepancy in article selection was resolved by consensus between the two reviewers.

Data extraction

Using the data extraction form in Table 3, YT recorded the relevant information in each included study, including language, research design, number of participants, age, country, definition of hearing loss, degree of hearing loss, type of hearing loss, age of hearing loss identification, use and type of hearing technology, acoustic environment, outcome measures,

audiovisual effects, additional results, level of evidence and quality rating, and other relevant information.

Table 3

Table Shell

Data items	Research design	Sample size	Age	Language	Country	Definition of hearing loss	Degree of hearing loss	Type of hearing loss	Age of hearing loss identification	Use and type of hearing technology	Acoustic environment	Outcome measures	Audiovisual effects	Results	Level of evidence & quality ratings
Definition	Any (e.g., quasi-experimental design, cross-sectional, etc.)	Any	Any; reported in years	Any	Any	Hearing loss is defined as hearing thresholds across five octave frequencies (0.25, 0.5, 1, 2, and 4 kHz) greater than 15 dB HL.	Any degree	Any type, including sensorineural, conductive, and mixed hearing loss	Any age, including congenital, pre-lingual, post-lingual acquired hearing loss	Unaided (no amplification) Hearing aid Cochlear Implant	Any SNR = Signal-to-noise ratio	Speech perception in quiet and in noise, including discrimination and recognition tasks for phoneme, tone, word, and sentence materials	Any		1-6 based on Cox (2005)

Assessment of level of evidence and quality of individual studies

Two reviewers (YT & NM) assessed the level of the evidence and the quality of each included study using the level of evidence hierarchy and the system of quality rating in Cox (2005). The level of evidence hierarchy (see Appendix B) consists of six levels, from level 1, the highest level evidence, "systematic reviews and meta-analyses of randomized controlled trials or other high-quality studies" to level 6, the lowest level evidence, expert opinion (Cox, 2005). The system of quality ratings indicates different levels of risk of bias of individual studies, and is used in conjunction with evidence Levels 1, 2, 3, and 4. Any discrepancy in level of evidence and quality ratings was resolved by consensus between the two reviewers.

Data analysis

The audiovisual effects on speech perception (including benefits, insignificant, and adverse effects) in each language were presented in a table and narrative synthesis.

Assessment of confidence in cumulative evidence

The two reviewers (YT & NM) assessed the cumulative evidence using the system of grading as recommended in Cox (2005) and assigned a grade (see Appendix C). The system for grading a recommendation comprises six grades, from grade A, "Level 1 or Level 2 studies with consistent conclusions", to grade D, "Level 6 evidence or inconsistent or inconclusive studies of any level or any studies that have a high risk of bias." Any discrepancy in grading the cumulative evidence was resolved by consensus between the two reviewers.

Results

The assessment of level of evidence and quality ratings, audiovisual effects, and assessment of confidence in cumulative evidence are presented groupings of languages and hearing technology.

The numbers of studies reviewed at each stage are displayed in *Figure 1*, illustrating the PRISMA flow diagram of the stages of this systematic review. First, at the identification stage, 2,907 references were identified through the database searching. Next, 1,672 duplicate references from across the different databases were automatically or manually removed. In the screening stage, 1,235 titles and abstracts were reviewed using the eligibility criteria. Of these, 1,176 studies were excluded as they were not journal articles or did not meet the eligibility criteria. At the eligibility stage, 59 references were reviewed and assessed using the full-text of the articles and 33 studies that did not meet the eligibility criteria were excluded. The reasons of exclusion of full-text articles included: the lack of participants with hearing loss, the lack of a speech perception measure, the lack of speech perception testing under an audiovisual condition, the lack of speech perception under either auditory-only or visual-only comparison conditions, non-English or non-Mandarin-speaking participants, and stimuli containing non-English and non-Mandarin phonemes. Finally, 26 studies were included in the systematic review.

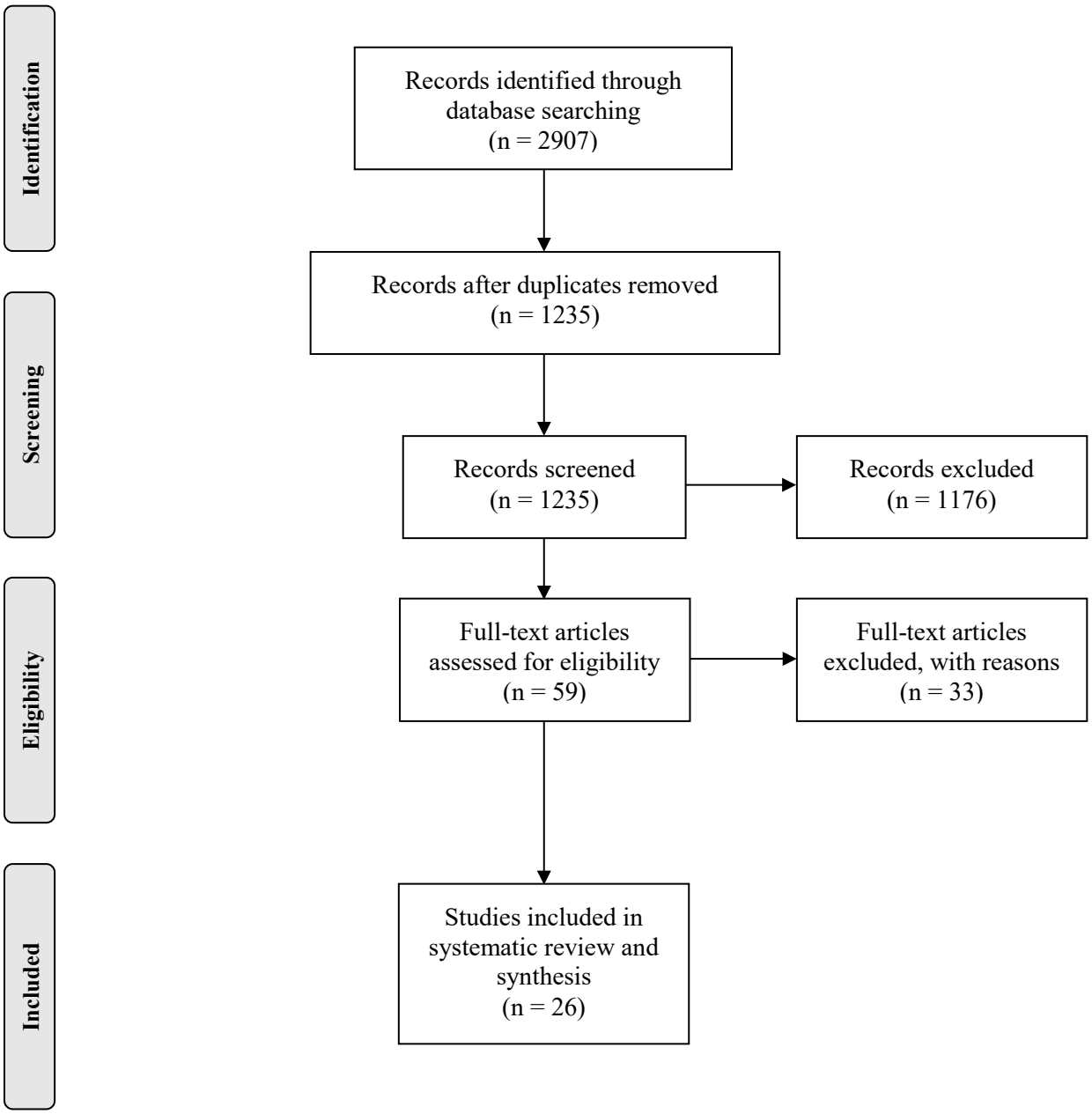

Figure 1. PRISMA flow diagram

Study design and characteristics of the included studies are presented in Table 4. Among the 26 studies, two of the studies were nonrandomized controlled trials, 11 were uncontrolled trials, two were cohort studies, one was cross-sectional study, one was case-control study, two were case series, and seven were case reports. With regard to language, 24 of the 26 included studies were related to English and two were related to Mandarin audiovisual speech perception. Of the 24 studies related to English audiovisual speech perception, 20 studies were conducted in the United States, three were performed in Australia, and one was conducted in Canada. Of the two studies related to audiovisual effects on Mandarin speech perception, one was conducted in China and one was performed in Taiwan. Concerning the age of participants, 16 of the included studies had adult participants, nine had children as participants, and one study did not specify the age of the participants. As to acoustic environments, 18 of the included studies were conducted in quiet, one was conducted in quiet but the stimuli were degraded by an 8-channel sinewave cochlear implant simulator, and seven were performed in noise at different signal-to-noise ratios. In terms of use and type of hearing technology, cochlear implants were involved in 11 studies, hearing aids were involved in seven studies, and participants were tested unaided in 11 studies. Regarding stimulus type, a range of stimuli and scoring methods were used across studies, including: phonemes in syllable context, nonsense syllables, closed-set monosyllabic words, spondee words, open-set words, sentences using syllable or keyword scoring, and sentence comprehension.

Table 4

Study Design and Characteristics of Included Articles

Study	Language	Research design	Number of Participants	Age	Country	Definition of hearing loss	Degree of hearing loss	Type of hearing loss	Age of hearing loss identification	Use and type of hearing technology	Acoustic environment
Altieri and Hudock (2014)	American English	case report	5 (3 males; 2 males)	22-72	United States	average pure tone threshold ≥ 25 dB SPL	varying degree average low frequency[a]: 8-25 average high frequency[b]: 25-50	1 conductive; 4 SNHL	1 sudden-onset; 1 from surgery as a toddler; 1 noise-exposure; 1 age-related; 1 unaware of HL	unaided	in quiet and degraded by an 8-channel sinewave CI simulator
Bergeson, Pisoni, and Davis (2003)	English	case series	80	children	United States	not specified	profound	not specified	before 36 months	CI	in quiet
Bergeson, Pisoni, and Davis (2005)	English	case series	80	children	United States	not specified	profound	not specified	before 36 months	CI	in quiet
Busby, Tong, and Clark (1984)	Australian	case report	4 (2 males; 2 females)	13.5-14	Australia	not specified	average 90 dB HTL at 500Hz 105dB HTL at 1kHz 105dB HTL at 2kHz 110dB HTL at 4 kHz	SNHL	congenital	bilateral HA	in quiet with ambient noise below 45 dBA
Busby, Tong, and Clark (1988)	Australian	case report	4 (2 males; 2 females)	13.5-14	Australia	not specified	average 90 dB HTL at 500 Hz 105 dB HTL at 1 kHz 105 dB	not specified	congenital	bilateral HA	in quiet with ambient noise below 45 dBA

							HTL at 2 kHz 110 dB HTL at 4 kHz				
Danhauer, Erratt, and Edgerton (1986)	American English	case report	10 (5 CI; 5 HA)	CI: 36-65 (mean: 55.6) HA: 34-68 (mean: 53.8)	United States	not specified	CI: severe to profound HA: moderately severe to profound	SNHL	CI: post-lingual HA: not specified	5 CI; 5 HA	in quiet
Danhauer, Garnett, and Edgerton (1985)	American English	case-control	15 (12 males; 3 females)	55-65	United States	normal hearing: ≤20 dB HL[c]	sloping	bilateral SNHL	not specified	1st presentation : unaided 2nd presentation : HA aided	in quiet
Desai, Stickney, and Zeng (2008)	English	case report	8	66	United States	not specified	deafened	not specified	post-lingual	CI	in quiet
Dorman et al. (2016)	English	uncontrolled trial	10 (4 males; 6 females)	21-87 (mean: 64)	United States	not specified	deafness	not specified	post-lingual	CI	in noise (multi-talker babble at an SNR of +3 - +10 dB that yielded approx. 40% correct responses in A for easy lists)
Erber (1979)	English	uncontrolled trial	Study 2: 2 Study 3: 12	Study 2: 12, 13 Study 3: 9-13	United States	not specified	Study 2: severe: 92 dB & 100 dB[d] Study 3: severe group (n = 6): 80-102 dB[d]	not specified	not specified	unaided	in quiet

							profound group (n = 6): mean 101-115 dB				
Goh, Pisoni, Kirk, and Remez (2001)	English	case report	1	35	United States	not specified	profound	not specified	29	CI	in quiet
Grant et al. (1998)	American English	uncontrolled trial	29	41-88 (mean: 65)	United States	not specified	33 dB HL[d] 53.5 dB HL[e]	SNHL	acquired (primarily noise exposure)	unaided	in speech-shaped noise at 0 dB SNR
Hack and Erber (1982)	American English	uncontrolled trial	18 (6 with good word recognition; 6 with intermediate word recognition; 6 with poor word recognition)	12:7-15:10	United States	not specified	83-123 dB	not specified	pre-lingual	unaided	in quiet (stimuli: 132-143 dB SPL ambient tape noise: 97 dB SPL)
Hay-Mccutcheon, Pisoni, and Hunt (2009)	English	cohort	25 (12 middle-aged; 13 elderly)	middle-aged: 41-54 (mean: 47) elderly: 66-81 (mean: 73)	United States	normal hearing: behavioral thresholds ≤25 dB HL[c]	not specified	not specified	2-69 middle-aged: 2-44 (mean: 17.3) elderly: 11-63 (mean: 41)	CI	in quiet
Hay-McCutcheon, Pisoni, and Kirk (2005)	English	cohort	34 (17 younger; 17 elderly)	younger adults: 39-53 (mean: 46) elderly adults: 65-83 (mean: 74)	United States	not specified	not specified	not specified	not specified	CI	in quiet

Holmes, Groccia, Johnson, and Green (1980)	English	uncontrolled trial	184 (99 males; 85 females)	6-15	United States	not specified	average 3-frequency: 95 dB	not specified	not specified	unaided	in quiet
Kirk et al. (2007)	American English	uncontrolled trial	15	children	United States	not specified	4 had a moderate or severe HL in the non-implanted ear; others had a bilateral profound HL	not specified	congenital or pre-lingual	CI	in quiet
Lei, Fang, Wang, and Mei (2008)	Mandarin Chinese	uncontrolled trial	19 (7 males; 12 females)	16.23	China	not specified	85.58 dB	not specified	not specified	HA	in quiet
Liu et al. (2014)	Mandarin Chinese	nonrandomized controlled trial	13 (8 males & 5 females)	18.1-56.5 (mean: 29.1 ± 13.5)	Taiwan	normal hearing: <25 dB HL[f]	severe to profound	bilateral SNHL	7 pre-lingually (before age 5); 6 post-lingually (after age 5)	CI	in quiet
Miller et al. (2017)	English	uncontrolled trial	76 (23 males; 53 females)	69	United States	mild to moderately severe[g]	mild to moderately-severe	bilateral SNHL	not specified	unaided	in noise (SSN & ISTS at an SNR of +8 dB)
Nicholson, Baum, Cuddy, and Munhall (2002)	English	case report	1	71	Canada	not specified	0.5 kHz: 25 dB HL 0.75 kHz: 35 dB HL 1 kHz: 45 dB HL 2 kHz: 50 dB HL 4 kHz: 75 dB HL	not specified	not specified	unaided	in noise (multi-voice speech babble)

Siegenthaler and Gruber (1969)	English	uncontrolled trial	32	adult	United States	not specified	not specified	not specified	not specified	HA tested aided & unaided	in quiet
Tye-Murray, Sommers, and Spehar (2007)	American English	nonrandomized controlled trial	26 (8 males; 18 females)	74.1	United States	normal hearing: PTA[d] < 25 dB HL	mild to moderate	SNHL	not specified	unaided	in noise (6-talker babble individually adjusted to achieve approx. 50% correct in A)
van Hoesel (2015)	Australian English	uncontrolled trial	7	not specified	Australia	not specified	not specified	not specified	not specified	bilateral CI	in noise
Walden, Busacco, and Montgomery (1993)	English	uncontrolled trial	40 (males) (20 middle-aged; 20 elderly)	middle-aged: 35-50 (mean: 42.5) elderly: 65-80 (mean: 72.2)	United States	not specified	moderate to severe	SNHL	not specified	HA	in noise (speech envelope noise): SNR that yields 40-50% correct recognition
Walden, Prosek, and Worthington (1974)	American English	cross-sectional	100	19-60	United States	not specified	not specified	not specified	not specified	unaided	in quiet

Note. SNHL = sensorineural hearing loss; CI = cochlear implant; HA = hearing aid; SNR = signal-to-noise ratio.
[a]250 Hz, 500 Hz, and 1000 Hz. [b]2000 Hz, 4000 Hz, and 8000 Hz. [c]From 250 Hz to 4000 Hz at octave frequencies. [d]Mean thresholds of 500 Hz, 1000 Hz, and 2000 Hz. [e]Mean threshold of 2000 and 4000 Hz. [f]At 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz. [g]No greater than 70 dB HL from 250 Hz through 4000 Hz.

Assessment of level of evidence and quality of individual studies

Each included study was assessed for level of evidence and quality. The level of evidence and quality rating of the studies included in this systematic review ranged from Level 3++ to Level 5. No randomized controlled trials were available. The level of evidence and quality rating of each study is presented in the results of included studies in Table 5 and is grouped by language and hearing technology in the following paragraphs.

Audiovisual effects in English-speaking cochlear implant users. The level of evidence and the quality of the studies related to audiovisual effects in English-speaking people with cochlear implants ranged from Level 3- to Level 5. For example, studies rated as Level 3- included Kirk et al. (2007), an uncontrolled trial with small sample size. Dorman et al. (2016) and van Hoesel (2015) were rated Level 3- as uncontrolled trials with small sample size and without recruitment methods reported. Hay-Mccutcheon et al. (2009) and Hay-McCutcheon et al. (2005), cohort studies with small sample size and without significant difference tested, were rated as Level 4-. The studies that were case reports or case series were rated as level 5, including Bergeson et al. (2003), Bergeson et al. (2005), Danhauer et al. (1986), Desai et al. (2008), and Goh et al. (2001).

Among the 26 studies included in this systematic review, 10 studies were associated with audiovisual speech perception in people with cochlear implants who speak English. Three of the 10 studies were rated as Level 3-, two were rated as Level 4-, and five were rated as Level 5.

Audiovisual effects in English-speaking hearing aid users. The level of evidence and quality in the studies relevant to audiovisual speech perception in English-speaking hearing aid users ranged from Level 3- to 5. Siegenthaler and Gruber (1969), an uncontrolled trial without testing significant differences and Walden et al. (1993), an uncontrolled trial with inadequate

sampling and without testing significant differences, were rated as Level 3-. To be more specific, the participants in Walden et al. (1993) were 40 male adults selected from Walter Reed Army Medical Center, which may not be sufficiently representative of adult population with hearing loss. Danhauer et al. (1985), a case control with small sample size and without reporting the recruitment method, was rated as Level 4-. Busby et al. (1984), Busby et al. (1988), and Danhauer et al. (1986) were rated as Level 5 due to being case reports.

Of the six studies related to audiovisual effects in people with hearing aids who speak English, two studies were rated as Level 3-, one was rated as Level 4-, and two were rated as Level 5.

Audiovisual effects in unaided English-speaking people with hearing loss. The level of evidence and quality of the studies associated with audiovisual speech perception in unaided English-speakers with hearing loss ranged from Level 3- to 5. Miller et al. (2017), an uncontrolled trial without weaknesses that may change the conclusions of the study, was rated as Level 3++. Erber (1979), Grant et al. (1998), Hack and Erber (1982), Holmes et al. (1980), Siegenthaler and Gruber (1969), and Tye-Murray et al. (2007) were rated as Level 3- on account of being uncontrolled trials or nonrandomized controlled trials with small sample size, inadequate sampling, and/or significant differences not tested. Danhauer et al. (1985), a case-control study with small sample size and without reporting recruitment methods, and Walden et al. (1974), a cross-sectional study without recruitment methods reported and without significant differences tested, were rated as Level 4-. Altieri and Hudock (2014) and Nicholson et al. (2002) were case reports and were rated as Level 5.

Among the 11 studies related to audiovisual effects in unaided English-speaking people with hearing loss, 1 study was rated as Level 3++, 6 studies were rated as Level 3-, 2 studies were rated as Level 4-, and 2 studies were rated as Level 5.

Audiovisual effects in Mandarin-speaking people with cochlear implants. Only one study, i.e. Liu et al. (2014), in this systematic review was related to audiovisual effects among Mandarin-speaking people with cochlear implants. The study was rated as Level 3- due to being a nonrandomized controlled trial with small sample size.

Audiovisual benefits in Mandarin-speaking people with hearing aids. Lei et al. (2008) was the only one study related to audiovisual speech perception in Mandarin-speaking hearing aid users in this systematic review. As an uncontrolled trial with small sample size, Lei et al. (2008) was rated as Level 3-.

Data synthesis

The full results of the included studies are presented in Table 5. Summaries of audiovisual benefits grouped by languages and hearing technology are presented in Table 6, Table 7, Table 8, and Table 9. Among the 26 included studies, audiovisual effects were statistically analyzed in 11 studies. Audiovisual benefits have been observed among English-speaking people with hearing loss regardless of age, degree of hearing loss, and use and type of hearing technology for consonant, vowel, syllable, word, and sentence recognition. Among Mandarin-speaking people with hearing loss, audiovisual benefits were found for phoneme and word recognition.

Table 5

Results of Included Studies

Language	Hearing technology	Study	Research design	Participants	Acoustic environment	Outcome measures	Main findings	Level of evidence & quality rating
English	unaided	Altieri and Hudock (2014)	case report	5 (3 male & 2 male) adults	in quiet and degraded by an 8-channel sinewave CI simulator	Exp 1: 75 sentences (25 AV, 25 A, 25 V) from the CUNY databases (open-set sentence recognition) and degraded using the CI simulator Exp 2: monosyllabic words "mouse", "job", "tile", "gain", "shop", "boat", "page", and "date" from the Hoosier Multi-Talker Database (closed-set speeded word recognition) and degraded using the CI simulator	Exp1: average scores AV > A average scores AV > V (significant differences not tested) Exp 2: average scores AV = A average scores AV > V (significant differences not tested)	5
English	CI	Bergeson, Pisoni, and Davis (2003)	case series	80 children	in quiet	PSI (close-set word & sentence recognition)	AV > A; AV > V (significant main effect of presentation format, $p < .0001$; post-hoc comparison not tested)	5
English	CI	Bergeson, Pisoni, and Davis (2005)	case series	80 children	in quiet	CP (open-set sentence comprehension)	AV > A; AV > V (significant main effect of presentation format, $p < .0001$; post-hoc comparison not tested)	5
English	HA	Busby, Tong, and Clark (1984)	case report	4 (2 male & 2 female children	in quiet with ambient noise below 45 dBA	11 Australian vowels in /h/-V-/d/ context (closed-set vowel identification)	average scores AV > A; average scores AV > V (significant differences not tested)	5
English	HA	Busby, Tong, and Clark (1988)	case report	4 (2 male & 2 female) children	in quiet with ambient noise below 45 dBA	14 consonants /p/, /t/, /k/, /b/, /d/, /g/, /f/, /v/, /z/, /ʒ/, /l/, and /r/ in /a/-C-/a/ context (closed-set	average scores AV1 > A; average scores AV1 > V (significant differences not tested)	5

English	CI & HA	Danhauer, Erratt, and Edgerton (1986)	case report	10 (5 CI & 5 HA) adults	in quiet	consonant identification) nonsense syllables (24 English consonants & 10 English vowels) & CID sentences (consonant, vowel, and sentence identification)	significant main effect of condition ($p < .01$) CI group: significant differences ($p < .01$) between AV & A, AV & V for consonant and sentence stimuli; significant differences between AV & A for vowel stimuli HA group: significant differences ($p < .01$) between AV & A and AV & V for consonant, vowel, and sentence stimuli	5
English	unaided & HA	Danhauer, Garnett, and Edgerton (1985)	case control	15 (12 male & 3 female) adults	in quiet	Nonsense Syllable Test (NST) (open-set phoneme recognition)	significant main effect of mode ($p < .001$) unaided: AV1 > V1 ($p < .01$); AV1 > A1 ($p < .01$) aided: AV2 > V2 ($p < .01$); AV2 > A2 ($p < .01$)	4- (small sample size; methods of recruiting participants not reported)
English	CI	Desai, Stickney, and Zeng (2008)	case report	8 adults	in quiet	11 CV in /ba/, /da/, and /ga/ continuum	average scores AV > A; AV > V (significant main effect of modality, $p < .05$; post-hoc comparison not tested)	5
English	CI	Dorman et al. (2016)	uncontrolled trial	10 (4 male & 6 female) adults	in noise (multi-talker babble at an SNR of +3 - +10 dB)	Kopra sentences (open-set sentence recognition scored by words correct)	mean AV > A for the difficult and east lists (significant effect of test modes, $p < .0001$)	3- (small sample size; methods of recruiting participants not reported)
English	unaided	Erber (1979)	uncontrolled trial	Study 2: 2 children Study 3: 12 children	in quiet	Study 2: 100 spondee words (open-set word recognition) Study 3: 80 sentences from the Magner Speech Intelligibility Test (open-set sentence recognition scored by content words correct)	Study 2: percent correct AV > V in both participants (significant differences not tested) Study 3: mean percent correct AV > V in both groups (significant differences not tested)	3- (small sample size; methods of recruiting participants for Study 3 not reported; significant differences not tested)
English	CI	Goh, Pisoni, Kirk, and Remez (2001)	case report	1 adult	in quiet	Background: CUNY Sentences Test (open-set sentence recognition) Exp: sinewave speech of 18 sentences from Remez et al.	Background: AV (100%) > A (92%); AV (100%) > V (63%) Exp: AV (89.7%) > A (52.5%); AV (89.7%) > V (43.2%)	5

English	unaided	Grant, Walden, and Seitz (1998)	uncontrolled trial	29 adults	in speech-shaped noise at 0 dB SNR	(1998) (sentence recognition scored by syllables correct) 18 consonant in /a/-C-/a/ context (closed-set consonant recognition); IEEE/Harvard (1969) sentences (open-set sentence recognition scored by key words)	consonant recognition: mean scores AV > A (significant differences not tested); mean scores AV > V (significant differences not tested) sentence recognition: mean scores AV > A (significant differences not tested); mean scores AV > V (significant differences not tested)	3- (inadequate sampling; significant differences not tested)
English	unaided	Hack and Erber (1982)	uncontrolled trial	18 children (6 with good word recognition, 6 with intermediate word recognition, & 6 with poor word recognition)	(in quiet) stimuli: 132-143 dB SPL ambient tape noise: 97 dB SPL	10 vowels in /b/-V-/b/ context (vowel identification)	mean scores AV > V in all 3 groups (significant differences not tested)	3- (small sample size; significant differences not tested)
English	CI	Hay-Mccutcheon, Pisoni, and Hunt (2009)	cohort	25 (12 middle-aged & 13 elderly) adults	in quiet	CUNY sentence test (sentence recognition scored by words correct)	mean AV > A; AV > V for both middle-aged and elderly groups (significant differences not tested)	4- (small sample size; significant differences not tested)
English	CI	Hay-McCutcheon, Pisoni, and Kirk (2005)	cohort	34 (17 younger & 17 elderly) adults	in quiet	CUNY sentence test (sentence recognition scored by words correct)	younger adults: mean percent correct AV > A; AV > V for pre-implantation and post-implantation (significant differences not tested) elderly adults: mean percent correct AV > A; AV > V for pre-implantation and post-implantation (significant differences not tested)	4- (small sample size; significant differences not tested)
English	unaided	Holmes, Groccia, Johnson, and Green (1980)	uncontrolled trial	184 (99 male & 85 female) children	in quiet	Word Intelligibility by Picture Identification (WIPI) (closed-set word recognition)	mean AV ≈ V; AV > A (significant differences not tested)	3- (significant differences not tested)
English	CI	Kirk et al. (2007)	uncontrolled trial	15 children	in quiet	Audiovisual-Lexical Neighborhood	AV > A ($p < .05$); AV > V ($p < .0001$) (significant main effect of presentation format, $p < .0001$)	3- (small sample size)

English	unaided	Miller et al. (2017)	uncontrolled trial	76 (23 male & 53 female) adults	in noise (SSN & ISTS at an SNR of +8 dB)	Sentence Test (AV-LNST) (sentence recognition scored by keywords correct) Multi-Modal Lexical Sentence Test for Adults (MLST-A) (open-set sentence recognition)	AV > A (significant main effect of presentation format, $p < .0001$)	3++
English	unaided	Nicholson, Baum, Cuddy, and Munhall (2002)	case report	1 adult	in noise (multi-voice speech babble)	speech-in-noise task (closed-set sentence identification scored by keyword correct)	AV > A by 25% (significant differences were not tested)	5
English	HA & unaided	Siegenthaler and Gruber (1969)	uncontrolled trial	32 adults	in quiet	PB-50 word lists (open-set word recognition)	aided: AV > A + V by mean 30%; unaided: AV > A + V by mean 20% in all but 1 subject	3- (significant differences not tested)
English	unaided	Tye-Murray, Sommers, and Spehar (2007)	nonrandomized controlled trial	26 (8 male & 18 female) adults	in noise (6-talker babble individually adjusted to achieve approx. 50% correct in A)	The Children's Audiovisual Enhancement Test (CAVET) (open-set word recognition)	mean scores AV > A > V for easy and hard words (significant differences not tested)	3- (significant differences not tested)
English	CI	van Hoesel (2015)	uncontrolled trial	7	in noise	Bamford-Kowal-Bench (BKB) sentences (open-set sentence recognition scored by target words correct)	average SRT50 AV < A for listening binaurally (significant effect of mode, $p = .034$)	3- (small sample size; methods of recruiting participants not reported)
English	HA	Walden, Busacco, and Montgomery (1993)	uncontrolled trial	40 male (20 middle-aged & 20 elderly) adults	in noise (speech envelope noise): SNR that yields 40-50% correct recognition	20 English consonants in C-/a/ context (closed-set consonant recognition); 45 sentences from the Central Institute for the Deaf Revised Sentences (sentence recognition	mean scores AV > A, AV > V in both middle-aged and elderly groups for sentence recognition and consonant recognition (significant differences not tested)	3- (inadequate sampling; significant differences not tested)

						scored by key words)		
English	unaided	Walden, Prosek, and Worthington (1974)	cross-sectional	100 adults	in quiet	20 English consonants in C-/a/ context (closed-set consonant recognition)	AV > A (significant differences not tested)	4- (methods of recruiting participants not reported; significant differences not tested)
Mandarin Chinese	HA	Lei, Fang, Wang, and Mei (2008)	uncontrolled trial	19 (7 male & 12 female) children	in quiet	course materials that contained 18 phonemes: /b/, /z/, /t/, /ch/, /q/, /k/, /a/, /o/, /e/, /i/, /u/, /ù/, /ai/, /uo/, /ie/, /an/, /ong/, /ao/ (phoneme identification)	AV > V (p = .002); AV > A (p = .000) (significant main effect of presentation format, p = .000)	3- (small sample size)
Mandarin Chinese	CI	Liu et al. (2014)	nonrandomized controlled trial	13 (8 male & 5 female) adults	in quiet	Mandarin Monosyllabic Word Recognition Test (MMRT) (word, phoneme, and tone reocognition)	phoneme recognition: AV > A in the pre-lingual group at SDT (p = .016) & SRT (p = .016) tone recognition: no significant difference word recognition: AV > A in the pre-lingual group at SDT (p = .016)	3- (small sample size)

Note. AV = audiovisual; A = auditory-only; V = visual only; CI = cochlear implant; HA = hearing aid; SNR = signal-to-noise ratio; SDT = speech detection thresholds; SRT = speech recognition thresholds.

Audiovisual effects in English-speaking cochlear implant users. Among English-speaking children with cochlear implants, better performance was reported in audiovisual than in auditory-only and visual-only conditions for closed-set word recognition, closed-set sentence recognition, and open-set sentence comprehension in quiet (Bergeson et al., 2003, 2005). Bergeson et al. (2003) assessed the closed-set word recognition and sentence recognition in quiet in 80 children with cochlear implants using the Pediatric Speech Intelligibility (PSI) test. The stimuli were presented via live voice. The results showed that the children performed better in the audiovisual condition than in the auditory-only and visual-only conditions and revealed a significant main effect of presentation mode ($p < .0001$). Using the Common Phrases (CP) test with the stimuli presented via live voice, Bergeson et al. (2005) investigated open-set sentence comprehension in quiet in 80 pre-lingually deaf children with cochlear implants. Better performance was revealed in the audiovisual condition compared to the performance in the auditory-only and visual-only conditions. A significant main effect of presentation format was found ($p < .0001$). Kirk et al. (2007) assessed speech perception in quiet using the Audiovisual-Lexical Neighborhood Sentence Test (AV-LNST) in 15 children with cochlear implants. Each sentence included three key words. The stimuli were presented through speakers approximately at 65 dB SPL. The children performed significantly better in the audiovisual conditions than in the auditory-only condition ($p < .05$) and in the visual-only condition ($p < .0001$).

Audiovisual benefits were also reported in English-speaking adults with cochlear implants. Danhauer et al. (1986) assessed audiovisual speech perception in quiet in five adult cochlear implant users using the Central Institute for the Deaf's (CID) Everyday sentences and the nonsense syllables that consisted of English consonants and vowels. The stimuli were presented via a loudspeaker and measured at 70 dBA at the participants' left ears. The

participants scored significantly higher in the audiovisual condition than in the auditory-only condition ($p < .01$) and in the visual-only condition ($p < .01$) for consonant recognition and sentence recognition and scored significantly higher in the audiovisual condition than in the auditory-only condition ($p < .01$) for vowel recognition. Desai et al. (2008) utilized 11 CV syllables in the /ba/, /da/, and /ga/ continuum as stimuli to measure speech perception in quiet in eight adults with cochlear implants. Seven of the eight participants were presented with the stimuli via a direct audio input to the processors and one was presented with the stimuli through a speaker. The results showed better mean scores in the audiovisual than in the auditory-only and in the visual-only conditions and revealed a significant main effect of modality ($p < .05$). Dorman et al. (2016) assessed speech perception in noise among ten adults with cochlear implant using Kopra sentences. The speech stimuli were presented at 60 dB SPL with multi-talker babble at a signal-to-noise ratio (SNR) between +3 and +10 dB. Better mean scores were reported in the audiovisual condition for the difficult and easy sentence lists compared with the mean scores in the auditory-only condition. A significant main effect of test mode was revealed for the two conditions ($p < .0001$), indicating the mean score was significantly higher in the audiovisual condition than in the auditory-only condition. Using Bamford-Kowal-Bench (BKB) sentences as stimuli, van Hoesel (2015) measured SRT50 in noise among seven cochlear implant users under audiovisual and auditory-only conditions. The age of the participants was not specified. The stimuli were presented via loudspeakers at 65 dBA. The results showed a significant main effect of mode for the two conditions ($p = .034$), suggesting that the mean SRT50 in the audiovisual condition was significantly lower than the mean SRT 50 in the auditory-only condition when the participants were tested binaurally. Regardless of age, significant audiovisual benefits were observed among English-speaking cochlear implant users.

However, some of the studies did not test for statistical significance in the performance between the audiovisual and other conditions; the audiovisual benefits were often shown by the difference in the mean scores. Using sinewave speech of 18 sentences as stimuli, Goh et al. (2001) measured speech perception in an exceptionally good adult cochlear implant user in quiet. The stimuli were presented via computer speakers approximately at 95 dB SPL and scored by syllables correctly responded. The results showed better recognition in the audiovisual than in the auditory-only and in the visual-only conditions. Hay-Mccutcheon et al. (2009) conducted a cohort study and compared the sentence recognition performance between middle-aged and elderly cochlear implant users. The participants included 12 middle-aged and 13 elderly adults with cochlear implants. The City University of New York (CUNY) sentence test was used and the stimuli were presented via a sound field speaker at 0° azimuth. For both middle-aged and elderly groups, better mean scores were reported in the audiovisual conditions compared with those in the auditory-only and in the visual-only conditions. No significant difference was found between the middle-aged and elderly groups for audiovisual and auditory-only conditions. Hay-McCutcheon et al. (2005) assessed the speech perception using the CUNY sentence test under audiovisual, auditory-only, and visual-only conditions in younger adults and elderly adults with cochlear implants before and after cochlear implantation. Seventy-four adults with cochlear implants were included in the study; seventeen participants were in the younger group and 17 were in the elderly group. The responses were scored by words correct. Both the younger and the elderly groups had higher mean scores in the audiovisual conditions than in auditory-only condition and in visual-only condition prior to cochlear implantation and after cochlear implantation. However, the audiovisual gain was found to be greater in the elderly group than in the younger group.

Out of 10 studies relevant to audiovisual speech perception in English-speaking people with cochlear implants, audiovisual benefits were found for consonant recognition in one of the studies, for vowel recognition in one study, for syllable recognition in one study, for word recognition in one study, and for sentence recognition or sentence comprehension in nine studies. A summary of audiovisual benefits for English-speaking cochlear implant users is presented in Table 6.

Table 6

Audiovisual Benefits in English-Speaking Cochlear Implant Users

Study	Hearing Technology	Types of Stimuli: Are there AV benefits?					Main effect of condition? ($p < 0.05$)
		Consonant	Vowel	Syllable	Word[a]	Sentence[b]	
Bergeson et al. (2003)	CI				yes	yes	yes
Bergeson et al. (2005)	CI					yes	yes
Danhauer et al. (1986) (CI group)	CI	yes[c]	yes[c]			yes[d]	yes
Desai et al. (2008)	CI			yes			yes
Dorman et al. (2016)	CI					yes	yes
Goh et al. (2001)	CI					yes	
Hay-Mccutcheon et al. (2009)	CI					yes	
Hay-McCutcheon et al. (2005)	CI					yes	
Kirk et al. (2007)	CI					yes[d]	yes
van Hoesel (2015)	CI					yes	yes

Note. CI = cochlear implant.
[a]Monosyllabic words. [b]Sentences scored by syllables correct or by words correct or sentence comprehension. [c]Stimuli were nonsense syllables. Post-hoc tests were statistically significant for vowels and for consonants. [d]Post-hoc tests were statistically significant.

Audiovisual effects in English-speaking hearing aid users. Audiovisual benefits were shown among English-speaking children with hearing aids. Busby et al. (1984) measured closed-set vowel identification among four children with hearing aids under audiovisual, auditory-only, and visual-only conditions in quiet with ambient noise below 45 dBA. The stimuli were eleven Australian English vowels in the context of /h/-V-/d/. A better mean score for vowel identification was reported in the audiovisual condition compared with the mean scores in the auditory-only and in the visual-only conditions. Using 14 consonants in the context of /a/-C-/a/ as stimuli, Busby et al. (1988) assessed closed-set consonant identification in quiet with ambient noise below 45 dBA among four children with hearing aids. The stimuli were presented via live voice and were monitored at 65 dBA at the subjects' position. The results showed that the children with hearing aids performed better in the audiovisual than in the auditory-only and in the visual-only conditions for consonant identification.

Audiovisual benefits were also discovered in adult hearing aid users who speak English. Using the CID sentences and nonsense syllables that consisted of English consonants and vowels as stimuli, Danhauer et al. (1986) measured sentence, consonant, and vowel identification under audiovisual, auditory-only, and visual-only conditions among five adults with hearing aids. Significantly better performance was reported in the audiovisual conditions than in the auditory-only and in visual-only conditions for the consonant, vowel, and sentence stimuli ($p < .01$ in all cases). Danhauer et al. (1985) assessed open-set phoneme recognition under audiovisual, auditory-only, and visual-only conditions in 15 adult hearing aid users. The Nonsense Syllable Test (NST) was used and the participants were tested unaided and aided. The stimuli were presented as CVCV nonsense syllables through the speaker of a video monitor and were monitored at 70 dB SPL at the participants' ears. When the participants were aided, significantly

better performance was found in the audiovisual condition compared with the performance in the visual-only condition ($p < .01$) and in the auditory-only condition ($p < .01$). Using PB-50 word lists as stimuli, Siegenthaler and Gruber (1969) measured the open-set word recognition in quiet under audiovisual, auditory-only, and visual-only conditions among 32 adults with hearing loss. The scores in the audiovisual conditions were compared with the sums of the scores in the auditory-only and in the visual-only conditions. The results showed that the scores in the audiovisual conditions were greater than the sum of the scores in auditory-only and visual-only conditions among all the participants when assessed aided ($M = 30\%$). That is, there was an audiovisual benefit among all the participants when aided. Walden et al. (1993) investigated closed-set consonant recognition and sentence recognition in noise in 40 male adults with moderate to severe sensorineural hearing losses and wearing hearing aids. All the participants were new hearing aid users. Among the 40 adults, 20 were middle-aged and 20 were elderly. The stimuli for consonant recognition were 20 English consonants followed by /a/, and 45 sentences from the Central Institute for the Deaf Revised Sentences were the stimuli for sentence recognition. The sentence stimuli were scored by key words correct. The stimuli were presented monaurally via an ER-1 earphone at 30 dB SL with speech envelop noise at an SNR that yielded 40% to 50% correct response. The result showed better mean scores for consonant and sentence recognition in the audiovisual than in the auditory-only and visual-only conditions in both middle-aged and elderly group. In addition, greater improvements were shown for sentence recognition than for consonant recognition and a significant main effect of test material was reported ($p < .01$).

Among the six studies associated with audiovisual speech perception in English-speaking hearing aid users, audiovisual benefits were reported for phoneme recognition in one study, for

consonant recognition in three studies, for vowel recognition in two studies, for word recognition in one study, and for sentence recognition in two studies. A summary of audiovisual benefits for English-speaking hearing aid users is presented in Table 7.

Table 7

Audiovisual Benefits in English-Speaking Hearing Aid Users

Study	Hearing Technology	Types of Stimuli: Are there AV benefits?					Main effect of condition? ($p < 0.05$)
		Phoneme	Consonant	Vowel	Word[a]	Sentence[b]	
Busby et al. (1984)	HA			yes[c]			
Busby et al. (1988)	HA		yes[d]				
Danhauer et al. (1986) (HA group)	HA		yes[e]	yes[e]		yes[f]	yes
Danhauer et al. (1985) (2nd presentation: aided)	HA	yes[e]					yes
Siegenthaler and Gruber (1969) (tested aided)	HA				yes		
Walden et al. (1993)	HA		yes[g]			yes	

Note. HA = hearing aid.
[a]Monosyllabic words. [b]Sentences scored by words correct. [c]Stimuli were presented as /h/-V-/d/. [d]Stimuli were presented as /a/-C-/a/. [e]Stimuli were nonsense syllables. Post-hoc tests were statistically significant. [f]Post-hoc tests were statistically significant. [g]Stimuli were presented as C-/a/.

Audiovisual effects in unaided English-speaking people with hearing loss.

Audiovisual benefits were revealed in English-speaking children with hearing loss when tested unaided. Erber (1979) investigated audiovisual speech perception in quiet among children with severe to profound hearing loss. In Study 2, two children at age 12 and 13 with profound hearing loss were assessed for open-set word recognition using 100 spondee words. In Study 3, 12 children at the age between nine and 13 were divided into two groups, the severe and profound groups. Using 80 sentences from the Magner Speech Intelligibility Test as stimuli, the children participating in Study 3 were assessed for recognition and the responses were scored by content words correct. The results of Study 2 showed better performance in the audiovisual conditions than in the visual-only conditions for spondee word recognition in both participants. With regard to Study 3, the results demonstrated higher mean scores in the audiovisual conditions compared with the mean scores in the visual-only conditions for the sentence stimuli in both severe and profound group. Besides, the contribution of the additional auditory cues to lipreading was shown to be greater in the severe group than in the profound group.

Audiovisual benefits were also found in English-speaking adults with hearing loss when tested unaided. Using the NST, Danhauer et al. (1985) assessed the open-set phoneme recognition among 15 adults with hearing loss in audiovisual, auditory-only, and visual-only conditions. The stimuli were CVCV nonsense syllables presented through the speaker of a video monitor and monitored at 70 dB SPL at the participants' ear. When tested unaided, the participants performed significantly better in the audiovisual condition than in the visual-only condition ($p < .01$) and in the auditory-only condition ($p < .01$). Miller et al. (2017) measured unaided sentence recognition in noise under audiovisual and auditory-only conditions among 76 adults with hearing loss using Multi-Modal Lexical Sentence Test for Adults (MLST-A). The

participants had bilateral sensorineural hearing loss and the degree of hearing loss ranged from mild to moderately severe hearing loss. The stimuli were presented via a loudspeaker at 65 dB SPL with background noise at an SNR of +8 dB. The results revealed a significant main effect of presentation format for the two conditions ($p < .0001$), indicating the mean score for sentence recognition was significantly better in the audiovisual than in the auditory-only conditions. Nicholson et al. (2002) investigated closed-set sentence identification in noise in an adult with hearing loss and right hemisphere damage. The sentence stimuli were presented with multi-voice speech babble and were scored by key words correctly responded. The intensity levels of the stimuli and the multi-voice speech babble were not specified. The results showed a better score in audiovisual condition than in auditory-only condition (by 25%) and in visual-only condition. Altieri and Hudock (2014) investigated open-set sentence recognition and closed-set speeded word recognition in quiet in five adults with varying degree of hearing loss under audiovisual, auditory-only, and visual-only conditions. The participants included four adults with sensorineural hearing loss and one adult with conductive hearing loss. Seventy-five sentences from the CUNY databases were used for the open-set sentence recognition, and monosyllabic words from the Hoosier Multi-Talker Database were used for closed-set speeded word recognition. The auditory content of the speech materials was degraded by an 8-channel sinewave cochlear implant simulator to avoid the ceiling effect and the accuracy under CI-simulated condition had been reported to be significantly correlated with the accuracy under multi-talker babble condition among normal-hearing listeners (Bent, Buchwald, & Pisoni, 2009). The auditory stimuli for speeded word recognition were presented via Beyer Dynamic-100 headphones approximately at 70 dB SPL. The results of the study showed a better mean score in audiovisual condition than in auditory-only condition and visual-only condition for the open-set

sentence recognition. As for the closed-set speeded word recognition, the mean score was better in the audiovisual condition than in the visual-only condition, whereas it was equivalent to the mean score in the auditory-only condition. Grant et al. (1998) assessed consonant recognition in the context of /a/-C-/a/ and open-set sentence recognition in noise among 29 adults with primarily noise-induced hearing loss under audiovisual, auditory-only, and visual-only conditions. The stimuli were presented binaurally via Beyer DT-770 headphones approximately at 85 dB SPL with speech-shaped noise at an SNR of 0 dB. Better mean scores were reported in audiovisual condition than in auditory-only condition and visual-only condition for consonant recognition and sentence recognition. For consonant recognition, significant correlations were shown between audiovisual and auditory-only scores ($r = .82$, $p < .0001$) and between audiovisual and visual-only scores ($r = .63$, $p < .001$). For sentence recognition, significant correlations were revealed between audiovisual and auditory-only scores ($r = .82$, $p < .001$) and between audiovisual and visual-only scores ($r = .44$, $p < .02$). Hack and Erber (1982) measured vowel identification in the context of /b/-V-/b/ in quiet under audiovisual, auditory-only, and visual-only conditions among 18 children with hearing loss. The children were divided into three groups based on their word recognition performance. Among the 18 children, six were identified with good word recognition, six with intermediated word recognition, and 6 with poor word recognition. The stimuli were presented via live voice and were presented monaurally via TDH-49 earphones at a comfortable listening level for each subject. Better mean scores were shown in the audiovisual condition than in the visual-only conditions in all three group with different levels of word recognition performance. However, the improvements were greater in children with good and intermediate word recognition than in children with poor recognition. Walden et al. (1974) measured consonant recognition in the context of C-/a/ in quiet under audiovisual,

auditory-only and visual-only conditions among 100 adults with hearing loss. The stimuli were presented monaurally via a TDH-49 earphone at 40 dB SL (re: SRT). Better scores were reported in the audiovisual condition than in the auditory-only condition. Using Word Intelligibility by Picture Identification (WIPI), Holmes et al. (1980) assessed closed-set word recognition in quiet among 184 deaf children under audiovisual, auditory-only, and visual-only conditions. The stimuli were presented through an FM auditory trainer with TDH-39 earphones and the level of the stimuli was adjusted to a comfortable listening level for each participant. The results showed better mean scores in the audiovisual condition than in auditory-only the condition. However, the mean scores in the visual-only condition approximated the mean scores in audiovisual conditions. It was also discovered that visual-only performance improved with age until the age of 12. Siegenthaler and Gruber (1969) assessed the open-set word recognition in quiet under audiovisual, auditory-only, and visual-only conditions among 32 adults with hearing loss using the PB-50 word lists. All except for one participants had scores in the audiovisual conditions greater than the sums of the scores in auditory-only and visual-only conditions when assessed unaided (M = 20%). Using the Children's Audiovisual Enhancement Test (CAVET), Tye-Murray et al. (2007) measured open-set word recognition in noise among 26 elderly adults with hearing loss in audiovisual, auditory-only, and visual-only conditions. The participants had sensorineural hearing loss and the degrees of hearing loss ranged from mild to moderate loss. The words were categorized as visually hard words or easy words based on the visibility. The stimuli were presented at 60 dB SPL with 6-talker babble in which the intensity levels were individually adjusted to yield approximately 50% correct responses in the auditory-only conditions to avoid the ceiling effect in the audiovisual conditions. The results showed that the mean score was greatest in the audiovisual condition, followed by the mean score in the auditory-only condition,

and with the lowest mean score in the visual-only condition for both easy and hard words among the elderly adults with sensorineural hearing loss.

Among the 11 studies related to audiovisual speech perception in unaided English-speaking people with hearing loss, audiovisual benefits were reported for phoneme recognition in one study, for consonant recognition in two studies, for vowel recognition in one study, for word recognition in five studies, and for sentence recognition in five studies. A summary of audiovisual benefits for unaided English-speaking people with hearing loss is presented in Table 8.

Table 8

Audiovisual Benefits in unaided English-Speaking People with Hearing Loss

Study	Hearing Technology	Types of Stimuli: Are there AV benefits?					Main effect of condition? ($p < 0.05$)
		Phoneme	Consonant	Vowel	Word[a]	Sentence[b]	
Altieri and Hudock (2014)	unaided				yes vs. V no vs. A	yes	
Danhauer et al. (1985) (1st presentation: unaided)	unaided	yes[c]					yes
Erber (1979)	unaided				yes	yes	
Grant et al. (1998)	unaided		yes[d]			yes	
Hack and Erber (1982)	unaided			yes[e]			
Holmes et al. (1980)	unaided				yes vs. A no vs.V		
Miller et al. (2017)	unaided					yes	yes
Nicholson et al. (2002)	unaided					yes	
Siegenthaler and Gruber (1969) (tested unaided)	unaided				yes		
Tye-Murray et al. (2007)	unaided				yes		
Walden et al. (1974)	unaided		yes[f]				

Note. V = visual-only; A = auditory-only.

[a]Monosyllabic words or spondee words.[b]Sentences scored by words correct. [c]Stimuli were nonsense syllables. Post-hoc tests were statistically significant. [d]Stimuli were presented as /a/-C-/a/. [e]Stimuli were presented as /b/-V-/b/. [e]Stimuli were sentences. [f]Stimuli were presented as C-/a/.

Audiovisual effects in Mandarin-speaking people with hearing loss. Liu et al. (2014) investigated audiovisual speech perception in quiet at different presentation levels under audiovisual and auditory-only conditions among 13 Mandarin-speaking adults with cochlear implants in Taiwan. Among the 13 participants with cochlear implants, seven were pre-lingually-deafened (prior to age five) and six were post-lingually deafened (after age five). The Mandarin Monosyllabic Word Recognition Test (MMRT) was used as stimuli and the participants were assessed for open-set word recognition, phoneme recognition, and tone recognition. In the seven pre-lingually deafened adults with cochlear implants, significant better scores were observed in the audiovisual conditions for phoneme recognition at speech detection thresholds (SDT) (p = .016) and at speech recognition thresholds (SRT) (p = .016) and for word recognition at SDT (p = .016). No significant differences were revealed at SRT+10. In the six post-lingually deafened adults using cochlear implants, no significant differences were found between the performance in the audiovisual condition and that in the auditory-only condition for phoneme, word, and tone recognition at any presentation levels. Neither pre-lingual nor post-lingual groups demonstrated significantly different performance between the audiovisual and auditory-only conditions for tone recognition. That is, additional visual information may facilitate phoneme and word recognition but not tone recognition in pre-lingually deafened adults with cochlear implants when the auditory signals are soft, and it may not help with speech perception in post-lingually deafened cochlear implant users and may not assist speech perception in Mandarin-speaking people with cochlear implants when the auditory signals are louder, i.e. SRT+10. The authors stated that the post-lingually deafened group had less reliance on visual information and that may be correlated with their language experiences prior to implantation and the automatic gain control of cochlear implants.

Audiovisual benefits were also shown in Mandarin-speaking children with hearing aids. Lei et al. (2008) assessed phoneme identification in quiet among 19 children with hearing aids in China. The stimuli were selected from the course materials that contained /b/, /z/, /t/, /ch/, /q/, /k/, /a/, /o/, /e/, /i/, /u/, /ü/, /ai/, /uo/, /ie/, /an/, /ong/, /ao/ and were presented via live voice by a trained female Mandarin speaker. The results showed that the children had the best phoneme identification performance in the audiovisual condition, followed by the performance in the visual-only condition, and had the worst performance in the auditory-only condition. A significant main effect of presentation format was revealed ($p = .000$). Post hoc analyses showed that the performance was significantly better in the audiovisual condition than in the auditory-only condition ($p = .000$) and in the visual-only condition ($p = .002$). In addition, the mean identification performance for the group of /a/, /o/, /e/, /i/, /u/, /ü/ was significantly better than the mean identification performance for the group of /b/, /z/, /t/, /ch/, /q/, /k/ ($p < .000$).

Audiovisual benefits for phoneme recognition were reported in both studies related to audiovisual speech perception in Mandarin-speaking people with hearing loss. However, it should be noted that significant audiovisual benefits for phoneme recognition were only found in children with hearing aids and pre-lingually deafened adults with cochlear implants but not in post-lingually deafened adult cochlear implant users. A summary of audiovisual benefits for Mandarin-speaking people with hearing loss is presented in Table 9.

Table 9

Audiovisual Benefits in Mandarin-Speaking People with Hearing Loss

Study	Hearing Technology	Types of Stimuli: Are there AV benefits?							Main effect of condition? ($p < 0.05$)
		Phoneme	Consonant	Vowel	Syllable	Tone	Word	Sentence	
Lei et al. (2008)	HA	yes							yes
Liu et al. (2014)	CI (pre-lingual)	yes[a, b] at SDT & at SRT; **no[a] at SRT+10**				**no**[a]	yes[b] at SDT; **no at SRT & at SRT +10**		
	CI (post-lingual)	**no**[a]				**no**[a]	**no**		

Note. HA = hearing aid; CI = cochlear implant; SDT = speech detection thresholds; SRT = speech recognition thresholds.
[a]Stimuli were words. [b]Post-hoc tests were statistically significant.

Assessment of confidence in cumulative evidence

Audiovisual effects in English-speaking cochlear implant users. As stated in Cox (2005), Grade B is assigned when the recommendation is supported by "consistent Level 3 or 4 studies or extrapolated evidence from Level 1 or 2 studies" and Grade C is assigned for a recommendation with "Level 5 studies or extrapolated evidence from Level 3 or 4 studies." As the evidence related to audiovisual effects in English-speaking cochlear implants users were composed of five studies of Level 3 or 4 and five Level 5 studies, Grade C+ was assigned.

Audiovisual effects in English-speaking hearing aid users. Supported by two Level 3- studies, one Level 4- studies, and three Level 5 studies showing audiovisual benefits, Grade C+ was assigned to the cumulative evidence of audiovisual benefits in English-speaking people with hearing.

Audiovisual effects in unaided English-speaking people with hearing loss. Comprised of mostly Level 3 and 4 studies with two Level 5 studies revealing audiovisual benefits, Grade B- was assigned to the cumulative evidence for audiovisual facilitation to speech perception in unaided English-speaking with hearing loss, indicating utilizing auditory in combination with visual information may help with speech perception in English-speaking people who have hearing loss and are unaided.

Audiovisual effects in Mandarin-speaking people with hearing loss. Among the 26 studies included in this systematic review, only two studies were related to audiovisual effects in Mandarin-speaking people with hearing loss. The age of the participants and the use of hearing technology was different in these two studies, i.e. adults with cochlear implants in Liu et al. (2014) and children with hearing aids in Lei et al. (2008). As a result, the assessment of

cumulative evidence was geared towards audiovisual effects in Mandarin-speaking people with hearing loss as a whole.

Since both Liu et al. (2014) and Lei et al. (2008) were rated as Level 3- and both of the studies showed audiovisual benefits for phoneme recognition, Grade B was assigned to the cumulative evidence of audiovisual benefits in Mandarin-speaking people with hearing loss, suggesting that audiovisual information may help with speech perception, specifically with phoneme recognition, among Mandarin-speaking people with hearing loss. Nonetheless, it should be noted that the cumulative evidence was assessed based on only two studies with small sample size. The results of the studies need to be interpreted with caution and future research with regard to audiovisual speech perception in the Mandarin-speaking population with hearing loss is warranted.

Discussion

Among the 26 studies that were included in this systematic review, 24 were related to audiovisual speech perception in English-speaking people with hearing loss and only two were related to audiovisual speech perception in Mandarin-speaking people with hearing loss. Apparently, there is a paucity of research on audiovisual effects on speech perception among people with hearing loss who speak Mandarin and future research in this area is warranted.

The limited numbers of and the heterogeneity across the studies related Mandarin audiovisual effects in people with hearing loss rendered the process of making recommendations of utilizing audiovisual information to facilitate speech perception in Mandarin challenging. Examining the only two studies on audiovisual speech perception of Mandarin-speaking people with hearing loss, the age of the participants, the type of hearing technology used, and the outcome measures used for assessing the speech perception, were different. Though the findings

of the studies were to some degree consistent, i.e. significant audiovisual benefits for phoneme recognition, caution needs to be applied when interpreting the findings due to the limited numbers of relevant studies available and the heterogeneity of the study characteristics.

With regard to the audiovisual effects on Mandarin speech perception among people with hearing loss, audiovisual benefits were revealed for phoneme and word recognition and the largest unit assessed in the studies was word. Consequently, it is not clear whether the audiovisual benefits at the word level can be expanded to a larger unit, e.g. sentence, and whether audiovisual benefits exist at a sentence level. In contrast, many studies have been conducted on the audiovisual effects on speech perception at a sentence level among English-speaking population with hearing loss and audiovisual benefits for English sentence stimuli have been revealed. Nonetheless, English and Mandarin are linguistically different (i.e. non-tonal vs. tonal) and recognition of a Mandarin sentence will involve recognition of the tones of the individual syllables in the sentence, and therefore the audiovisual benefits shown for English sentence recognition may not be found for Mandarin sentence recognition.

Examining all the included studies, the characteristics of the studies were heterogeneous, making the assessment of the cumulative evidence a challenge. To be more specific, the degree and type of the hearing loss the participants experienced varied from study to study. For example, the participants in Tye-Murray et al. (2007) had mild to moderate sensorineural hearing loss while Altieri and Hudock (2014) included people with sensorineural hearing loss as well as people with conductive hearing loss in their studies and the degree of hearing loss varied from participant to participant. Some of the studies even did not specify the type of hearing loss of the participants. In addition to the diverse characteristics of the participants, the outcome measures

used were inconsistent across studies, which renders a meta-analysis of the study findings not feasible.

Another challenge of interpretation of the findings was the differences in the manner stimuli were presented. Stimuli were presented through loudspeakers, headphones, or direct audio input and the levels of the intensity presented varied across studies. Some research presented stimuli and monitored the intensity level of the stimuli at a specific level for all the participants, while in other studies, the intensity levels of the stimuli were adjusted by the participants to their comfortable listening levels. That is, the intensity levels of the stimuli presented to each participant may be different.

Clinical Implications

Audiovisual information facilitates speech perception in English-speaking people with hearing loss regardless of age, degree of hearing loss, use and type of hearing technology, and acoustic environment. Audiovisual information also assists in speech perception among Mandarin-speaking people with hearing loss, depending on the unit of speech perception measured. However, the benefits were only found for phoneme recognition in children with hearing aids and in pre-lingually deafened adults with cochlear implants, and for word recognition in pre-lingually deafened adult cochlear implant users. No significant audiovisual benefits were shown for Mandarin tone recognition and for speech perception at higher intensity levels. No significant audiovisual benefits were discovered among post-lingually deafened adults with cochlear implants. Based on the cumulative evidence, it is recommended that people with hearing loss utilize audiovisual information to assist speech perception, but the recommendation should be interpreted with caution for Mandarin-speaking population with hearing loss due to the limited amount of studies and the heterogeneity of the study characteristics.

Conclusion

People with hearing loss who speak English or Mandarin are encouraged to use visual cues in addition to auditory information to facilitate speech perception. However, given the results of the systematic review, it appears that future research on audiovisual speech perception among Mandarin-speaking people with hearing loss is warranted and that consistent outcome measures for speech perception are needed for performing a meta-analysis on audiovisual speech perception among people with hearing loss. Audiovisual information for people with hearing loss across languages seems advantageous to speech perception in people with hearing loss with different ages, degrees of hearing loss, and use and types of hearing technology.

Appendix A – Search Strategies

Database: Embase

Controlled vocabulary: Emtree

Search date: 11/7/2018

Search strategy:

('hearing impairment'/exp OR 'hearing impairment' OR 'hearing loss' OR deaf OR deafness OR 'hearing impaired' OR 'hearing-impaired' OR 'hard of hearing')

AND ('speech perception'/exp OR 'speech perception')

AND (audiovisual OR 'audio-visual' OR 'auditory-visual' OR 'auditory visual' OR 'visual cues' OR 'visual benefits' OR 'visual contribution' OR 'lip reading'/exp OR 'lip reading' OR 'lip-reading' OR lipreading OR speechreading OR 'speech reading' OR 'speech-reading')

Limit: humans (Quick limits)

Database: PubMed

Controlled vocabulary: MeSH

Search date: 11/7/2018

Search strategy:

("Hearing Loss"[Mesh] OR "hearing loss" OR "hearing impairment" OR deaf OR deafness OR "hearing impaired" OR "hearing-impaired" OR "hard of hearing")

AND ("Speech Perception"[Mesh] OR "speech perception")

AND (audiovisual OR "audio-visual" OR "auditory-visual" OR "auditory visual" OR "visual cues" OR "visual benefits" OR "visual contribution" OR "Lipreading"[Mesh] OR lipreading OR "lip reading" OR "lip-reading" OR speechreading OR "speech reading" OR "speech-reading")

Limit: Humans (Species)

Database: Scopus

No controlled vocabulary

Search date: 11/7/2018

Search strategy:

TITLE-ABS-KEY ("hearing loss" OR "hearing impairment" OR deaf OR deafness OR "hearing impaired" OR "hearing-impaired" OR "hard of hearing")

AND TITLE-ABS-KEY ("speech perception")

AND TITLE-ABS-KEY (audiovisual OR "audio-visual" OR "auditory-visual" OR "auditory visual" OR "visual cues" OR "visual benefits" OR "visual contribution" OR lipreading OR "lip reading" OR "lip-reading" OR speechreading OR "speech reading" OR "speech-reading")

Limit: Human (Keyword)

Database: CINAHL Plus

Controlled vocabulary: CINAHL Heading

Search date: 11/7/2018

Search strategy:

((MH "Deafness") OR (MH "Hearing Loss, Partial+") OR "hearing loss" OR "hearing impairment" OR deaf OR deafness OR "hearing impaired" OR "hearing-impaired" OR "hard of hearing")

AND ((MH "Speech Perception") OR "speech perception")

AND (audiovisual OR "audio-visual" OR "auditory-visual" OR "auditory visual" OR "visual cues" OR "visual benefits" OR "visual contribution" OR (MH "Lipreading") OR lipreading OR "lip reading" OR "lip-reading" OR speechreading OR "speech reading" OR "speech-reading")

(did not explode “Deafness” in order not to include “Deaf-Blind Disorders”)

Limit: Human (Advanced Search)

Database: Web of Science

No controlled vocabulary

Search date: 11/7/2018

Search strategy:

TOPIC: ("hearing loss" OR "hearing impairment" OR deaf OR deafness OR "hearing impaired" OR "hearing-impaired" OR "hard of hearing")

AND TOPIC: ("speech perception")

AND TOPIC: (audiovisual OR "audio-visual" OR "auditory-visual" OR "auditory visual" OR "visual cues" OR "visual benefits" OR "visual contribution" OR lipreading OR "lip reading" OR "lip-reading" OR speechreading OR "speech reading" OR "speech-reading")

Database: PsycINFO

Controlled vocabulary: Thesaurus

Search date: 11/7/2018

Search strategy:

(DE "Deaf" OR DE "Partially Hearing Impaired" OR "hearing loss" OR "hearing impairment" OR deaf OR deafness OR "hearing impaired" OR "hearing-impaired" OR "hard of hearing")

AND (DE "Speech Perception" OR "speech perception")

AND (audiovisual OR "audio-visual" OR "auditory-visual" OR "auditory visual" OR "visual cues" OR "visual benefits" OR "visual contribution" OR DE "Lipreading" OR lipreading OR "lip reading" OR "lip-reading" OR speechreading OR "speech reading" OR "speech-reading")

Limit: Human (Population Group)

Appendix B – Level of Evidence Hierarchy and System of Quality Rating (Cox, 2005)

Level	Type of Evidence
1	Systematic reviews and meta-analyses of randomized controlled trials or other high-quality studies.
2	Randomized controlled trials
3	Nonrandomized intervention studies.
4	Nonintervention studies: cohort studies, case-control studies, cross-sectional surveys.
5	Case reports
6	Expert opinion.

Rating	Interpretation of Rating
++	Very low risk of bias. Any weaknesses that are present are very unlikely to alter the conclusions of the study
+	Low risk of bias. Identified weaknesses or omitted information probably would not alter the conclusions of the study.
-	High risk of bias. Identified weaknesses or omitted information are likely or very likely to alter the conclusions of the study.

Appendix C – System for Grading a Recommendation (Cox, 2005)

Grade	Criteria for grade assignment
A	Level 1 or Level 2 studies with consistent conclusions.
B	Consistent Level 3 or 4 studies or extrapolated evidence* from Level 1 or 2 studies.
C	Level 5 studies or extrapolated evidence from Level 3 or 4 studies.
D	Level 6 evidence or inconsistent or inconclusive studies of any level or any studies that have a high risk of bias.

References

Altieri, N., & Hudock, D. (2014). Hearing impairment and audiovisual speech integration ability: a case study report. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00678

American Speech-Language-Hearing Association. (n.d.-a). Adult audiologic (hearing) rehabilitation. Retrieved from https://www.asha.org/public/hearing/Adult-Aural-Rehabilitation/

American Speech-Language-Hearing Association. (n.d.-b). Child aural/audiologic rehabilitation. Retrieved from https://www.asha.org/public/hearing/treatment/child_aur_rehab.htm

American Speech-Language-Hearing Association. (n.d.-c). English phonemic inventory. Retrieved from https://www.asha.org/uploadedFiles/practice/multicultural/EnglishPhonemicInventory.pdf

American Speech-Language-Hearing Association. (n.d.-d). Mandarin phonemic inventory. Retrieved from https://www.asha.org/uploadedFiles/practice/multicultural/MandarinPhonemicInventory.pdf

Bent, T., Buchwald, A., & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *The Journal of the Acoustical Society of America, 126*(5), 2660-2669. doi:10.1121/1.3212930

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *Volta Review, 103*(4), 347-370.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear and Hearing, 26*(2), 149-164. doi:10.1097/00003446-200504000-00004

Busby, P. A., Tong, Y. C., & Clark, G. M. (1984). Underlying dimensions and individual differences in auditory, visual, and auditory-visual vowel perception by hearing-impaired children. *The Journal of the Acoustical Society of America, 75*(6), 1858-1865.

Busby, P. A., Tong, Y. C., & Clark, G. M. (1988). Underlying structure of auditory-visual consonant perception by hearing-impaired children and the influences of syllabic compression. *Journal of Speech and Hearing Research, 31*(2), 156-165.

Chen, T. H., & Massaro, D. W. (2011). Evaluation of synthetic and natural Mandarin visual speech: Initial consonants, single vowels, and syllables. *Speech Communication, 53*(7), 955-972. doi:https://doi.org/10.1016/j.specom.2011.03.009

Cox, R. M. (2005). Evidence-based practice in provision of amplification. *Journal of the American Academy of Audiology, 16*(7), 419-438.

Danhauer, J. L., Erratt, J. D., & Edgerton, B. J. (1986). Auditory/visual speech perception by cochlear implant users and hearing aid wearers. *The American Journal of Otology, 7*(5), 354-360.

Danhauer, J. L., Garnett, C. M., & Edgerton, B. J. (1985). Older persons' performance on auditory, visual, and auditory–-visual presentations of the edgerton and danhauer nonsense syllable test. *Ear and Hearing, 6*(4), 191-197. doi:10.1097/00003446-198507000-00004

Desai, S., Stickney, G., & Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America, 123*(1), 428-440. doi:10.1121/1.2816573

Dorman, M. F., Liss, J., Wang, S., Berisha, V., Ludwig, C., & Natale, S. C. (2016). Experiments on auditory-visual perception of sentences by users of unilateral, bimodal, and bilateral cochlear implants. *Journal of Speech, Language, and Hearing Research, 59*(6), 1505-1519. doi:10.1044/2016_JSLHR-H-15-0312

Erber, N. P. (1979). Speech perception by profoundly hearing-impaired children. *Journal of Speech and Hearing Disorders, 44*(3), 255-270.

Goh, W. D., Pisoni, D. B., Kirk, K. I., & Remez, R. E. (2001). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. *Ear and Hearing, 22*(5), 412-419.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103*(5), 2677-2690. doi:10.1121/1.422788

Hack, Z. C., & Erber, N. P. (1982). Auditory, visual, and auditory-visual perception of vowels by hearing-impaired children. *Journal of Speech and Hearing Research, 25*(1), 100-107.

Hay-Mccutcheon, M. J., Pisoni, D. B., & Hunt, K. K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *International Journal of Audiology, 48*(6), 321-333. doi:10.1080/14992020802644871

Hay-McCutcheon, M. J., Pisoni, D. B., & Kirk, K. I. (2005). Audiovisual speech perception in elderly cochlear implant recipients. *The Laryngoscope, 115*(10 I), 1887-1894. doi:10.1097/01.mlg.0000173197.94769.ba

Holmes, D. W., Groccia, B., Johnson, K., & Green, W. (1980). Deaf children's processing of auditory, visual, and combined stimuli. *Ear and Hearing, 1*(3), 126-129.

Jachimski, D., Czyzewski, A., & Ciszewski, T. (2018). A comparative study of English viseme recognition methods and algorithms. *Multimedia Tools and Applications, 77*(13), 16495-16532. doi:10.1007/s11042-017-5217-5

Kirk, K. I., Hay-McCutcheon, M. J., Holt, R. F., Gao, S., Qi, R., & Gerlain, B. L. (2007). Audiovisual spoken word recognition by children with cochlear implants. *Audiological Medicine, 5*(4), 250-261. doi:10.1080/16513860701673892

Lei, J., Fang, J., Wang, W., & Mei, Y. (2008). The visual-auditory effect on hearing-handicapped students in lip-reading Chinese phonetic identification. *Psychological Science, 31*(2), 312-314.

Li, H., & Tang, C. J. (2011). Dynamic Chinese viseme model based on phones and control function. *Electronics Letters, 47*(2), 144-145. doi:10.1049/el.2010.2570

Liu, S.-Y., Yu, G., Lee, L.-A., Liu, T.-C., Tsou, Y.-T., Lai, T.-J., & Wu, C.-M. (2014). Audiovisual speech perception at various presentation levels in Mandarin-speaking adults with cochlear implants. *PLOS ONE, 9*(9). doi:10.1371/journal.pone.0107252

Miller, C. W., Stewart, E. K., Wu, Y.-H., Bishop, C., Bentler, R. A., & Tremblay, K. (2017). Working memory and speech recognition in noise under ecologically relevant listening conditions: Effects of visual cues and noise type among adults with hearing loss. *Journal*

of Speech Language and Hearing Research, 60(8), 2310-2320. doi:10.1044/2017_jslhr-h-16-0284

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Nicholson, K. G., Baum, S., Cuddy, L. L., & Munhall, K. G. (2002). A case of impaired auditory and visual speech prosody perception after right hemisphere damage. *Neurocase, 8*(4), 314-322. doi:10.1093/neucas/8.4.314

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics, 59*(1), 73-80. doi:10.3758/BF03206849

Siegenthaler, B. M., & Gruber, V. (1969). Combining vision and audition for speech reception. *Journal of Speech and Hearing Disorders, 34*(1), 58-60.

Simons, G. F., & Fennig, C. D. (2018). Ethnologue: Languages of the world. Retrieved from http://www.ethnologue.com

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification, 11*(4), 233-241. doi:10.1177/1084713807307409

Tyler, R. S., Fryauf-Bertschy, H., Kelsay, D. M. R., Gantz, B. J., Woodworth, G. P., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear

implants. *Otolaryngology - Head and Neck Surgery, 117*(3), 180-187. doi:10.1016/S0194-5998(97)70172-4

United States Census Bureau. (2015). Detailed languages spoken at home and ability to speak English for the population 5 years and over for United States: 2009-2013. Retrieved from https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

van Hoesel, R. J. M. (2015). Audio-visual speech intelligibility benefits with bilateral cochlear implants when talker location varies. *Journal of the Association for Research in Otolaryngology, 16*(2), 309-315. doi:10.1007/s10162-014-0503-7

Walden, B. E., Busacco, D. A., & Montgomery, A. A. (1993). Benefit from visual cues in auditory, visual speech recognition by middle-aged and elderly persons. *Journal of Speech and Hearing Research, 36*(2), 431-436.

Walden, B. E., Prosek, R. A., & Worthington, D. W. (1974). Predicting audiovisual consonant recognition performance of hearing-impaired adults. *Journal of Speech and Hearing Research, 17*(2), 270-278. doi:10.1044/jshr.1702.270

World Health Organization. (2018). Deafness and hearing loss. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss